**Carlos Henrique Botelho**

**RA: 5536403**

**EXERCÍCIOS DA AULA 7B – REGRESSÃO LINEAR MULTIPLA**

**EXERCÍCIO 1 - Recem-Nascidos**

**babies <- read.table("babies.txt",header=T,na.strings=999)**

**str(babies)**

**summary(babies)**

**babies$age[babies$age==99] <- NA**

**babies$height[babies$height==99] <- NA**

**babies$smoke[babies$smoke==9] <- NA**

**babies$smoke <- as.logical(babies$smoke)**

**table(apply(is.na(babies),1,sum))**

**babies <- babies[apply(is.na(babies),1,sum)==0,]**

**summary(babies)**

**str(babies)**

**# Mod.1**

**plot(bwt~gestation, data=babies)**

**abline(h=mean(babies$bwt), lty=2, col="red")**

**mod.1<- lm(bwt~gestation, data=babies)**

**abline(mod.1, col="blue")**

**par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))**

**plot(mod.1)**

**coef(mod.1)**

**anova(mod.1)**

**summary(mod.1)**

**#Preditos**

**gestation<- babies$gestation**

**gestation**

**class(gestation)**

**range(gestation)**

**predict<-predict( mod.1, data.frame( gestation=c(148:353, length(100)) ) )**

**predict**

**## Coeficientes do modelo**

**cf.mod1 <- coef(mod.1)**

**cf.mod1**

**head(model.matrix(mod.1))**

**head(model.matrix(mod.1)%\*%cf.mod1)**

**tail(model.matrix(mod.1)%\*%cf.mod1)**

**plot( resid( mod.1 ), fitted( mod.1 ) )**

## EXERCÍCIO 2 - Galileu estava Certo?

# Ajuste de Polinômios

init.h <- c(600, 700, 800, 950, 1100, 1300, 1500)

h.d <- c(253, 337, 395, 451, 495, 534, 573)

plot(h.d~init.h)

mod1 <- lm(h.d~init.h)

mod2 <- update(mod1,.~. +I(init.h^2))

anova(mod1,mod2)

abline(mod1)

cf.m2 <- coef(mod2)

curve(cf.m2[1]+cf.m2[2]\*x+cf.m2[3]\*x^2, add=T, lty=2)

summary(mod2)

# Polinomio de grau 3

mod3 <- update(mod2,.~. +I(init.h^3))

# Teste para comparar 2 modelos( mod2 e mod3).

anova(mod2,mod3)

# O valor de p no teste ANOVA é estatisticamente significativo.

cf.m3 <- coef(mod3)

curve(cf.m3[1]+cf.m3[2]\*x+cf.m3[3]\*x^2+cf.m3[4]\*x^3, add=T, lty=2)

summary(mod3)

**# R: o polinômio de grau 3 representa melhor a relação entre hd e init.h do que o polinômio de grau 2 e 1.**

**EXERCÍCIO 3**

1º Plotar todas as relações 2 a 2 da variável dependente (resposta) em relação a independente (preditora) com o objetivo de verificar: Linearidade, Independência, Normalidade e Igualdade de variância.

Escolher a partir dos gráficos quais variáveis acrescentaremos no modelo.

# Exercicio 7b.3 - Recem-Nascidos

babies <- read.table("babies.txt",header=T,na.strings=999)

str(babies)

summary(babies)

## Corrigindo os outros dados com código de NA

##(9 ou 99, dependendo da variavel)

###############################################

babies$age[babies$age==99] <- NA

babies$height[babies$height==99] <- NA

babies$smoke[babies$smoke==9] <- NA

babies$smoke <- as.logical(babies$smoke)

## Eliminando linhas com algum dado faltante

table(apply(is.na(babies),1,sum))

babies <- babies[apply(is.na(babies),1,sum)==0,]

summary(babies)

str(babies)

###############################################

# Variaveis

par(mfrow=c(1,3))

plot(bwt~gestation, data=babies) # lm.babies1

plot(bwt~parity, data=babies) # lm.babies2

plot(bwt~age, data=babies) # lm.babies3

par(mfrow=c(1,1))

par(mfrow=c(1,3))

plot(bwt~height, data=babies) # lm.babies4

plot(bwt~weight, data=babies) # lm.babies5

plot(bwt~smoke, data=babies) # lm.babies6

par(mfrow=c(1,1))

# lm.babies1 - Incorporar no modelo, valor de p significativo.

# Multiple R-squared: 0.1661 e Adjusted R-squared: 0.1654 explica consideravelmente o bwt

plot(bwt~gestation, data=babies)

abline(h=mean(babies$bwt), lty=2, col="red")

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

lm.babies1 <- lm(bwt~gestation, data=babies)

abline(lm.babies1, col="blue")

plot(lm.babies1)

coef(lm.babies1)

anova(lm.babies1)

summary(lm.babies1)

# lm.babies2 - Incorporar no modelo por ultimo.

plot(bwt~parity, data=babies)

abline(h=mean(babies$bwt), lty=2, col="red")

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

lm.babies2 <- lm(bwt~parity, data=babies)

abline(lm.babies2, col="blue")

plot(lm.babies2)

coef(lm.babies2)

anova(lm.babies2)

summary(lm.babies2)

# lm.babies3 - Não incorporar, pois o valor de p nao e significativo.

plot(bwt~age, data=babies)

abline(h=mean(babies$bwt), lty=2, col="red")

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

lm.babies3 <- lm(bwt~age, data=babies)

abline(lm.babies3, col="blue")

plot(lm.babies3)

coef(lm.babies3)

anova(lm.babies3)

summary(lm.babies3)

# lm.babies4 - Nao incorporar, pois o R squared e baixo, o test t nao e sgnificativo

# a variavel independente nao ajuda a explicar o modelo. Angulo do coef e quase zero.

plot(bwt~height, data=babies)

abline(h=mean(babies$bwt), lty=2, col="red")

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

lm.babies4 <- lm(bwt~height, data=babies)

abline(lm.babies4, col="blue")

plot(lm.babies4)

coef(lm.babies4)

anova(lm.babies4)

summary(lm.babies4)

# lm.babies5 - Nao incorporar Multiple R-squared: 0.02431 e Adjusted R-squared: 0.02348 muito baixo.

plot(bwt~weight, data=babies)

abline(h=mean(babies$bwt), lty=2, col="red")

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

lm.babies5 <- lm(bwt~weight, data=babies)

abline(lm.babies3, col="blue")

plot(lm.babies5)

coef(lm.babies5)

anova(lm.babies5)

summary(lm.babies5)

# lm.babies6 - Nao incorporar Multiple R-squared: 0.06091 e Adjusted R-squared: 0.06011,

# valores sao baixo e não explicam o bwt.

plot(bwt~smoke, data=babies)

abline(h=mean(babies$bwt), lty=2, col="red")

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

lm.babies6 <- lm(bwt~smoke, data=babies)

abline(lm.babies6, col="blue")

plot(lm.babies6)

coef(lm.babies6)

anova(lm.babies6)

summary(lm.babies6)

# Variaveis incorporadas ao modelo

# Mod.1

plot(bwt~gestation, data=babies)

abline(h=mean(babies$bwt), lty=2, col="red")

mod.1<- lm(bwt~gestation, data=babies)

abline(lm.babies6, col="blue")

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

plot(mod.1)

coef(mod.1)

anova(mod.1)

summary(mod.1)

# Mod.2

mod.2<- lm(bwt~gestation+parity, data=babies)

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

plot(mod.2)

coef(mod.2)

anova(mod.1,mod.2)

summary(mod.2)

# Mod.3

mod.3<- lm(bwt~gestation+ I(gestation^2) + parity, data=babies)

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

plot(mod.3)

coef(mod.3)

anova(mod.2,mod.3)

summary(mod.3)

# Mod.4

mod.4<- lm(bwt~I(gestation^2) + parity, data=babies)

par(cex=1, cex.lab=1.3, cex.main= 2, las= 1, bty= "l", tcl=0.5, mar= c(5,5,5,3))

plot(mod.4)

coef(mod.4)

anova(mod.2,mod.4)

summary(mod.4)

# Mod.2>Mod.3>Mod.4

**# Acredito que o Mod.2 é a que melhor explica a relação bwt.**